U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 373

QUALITY CONTROL OF SIGNIFICANT LEVEL RAWINSONDE TEMPERATURES AND PRESSURES

WILLIAM G. COLLINS

DEVELOPMENT DIVISION

NOVEMBER 1990

Introduction.

The quality control of meteorological data by a methodology called Complex Quality Control (CQC) (Gandin, 1988), has been developed and implemented in parts as each becomes available at the National Meteorological Center since 1988. Two principles of CQC are important to understanding its development. First, it is specific to the observing platform and observed variable. Second, the method uses several, independent checks before making any decision on the quality of a datum. Another important aspect of CQC is that it is designed not only to identify bad data, but to correct data when possible. In order to correct data, it is necessary to determine the error source, and this is the feature that leads to the requirement that the method be observing system dependent.

The first element of CQC to be implemented at NMC was the hydrostatic checking of rawinsonde heights and temperatures. The hydrostatic check is not the only one that can be performed on these data. Horizontal and vertical statistical checks and others could also be performed. But the hydrostatic check is powerful by itself for identifying and correcting "human" errors (not responding at all to instrument errors), and can be performed on single profiles of heights and temperatures, allowing its use very soon after data decoding.

The significant level temperatures are used within the Regional Data Assimilation System and will soon be used in the so-called Unified Global Data Assimilation System. And yet these data are not checked manually and not very rigorously by the assimilation systems. For this reason, it was natural that an automatic method be developed for checking these data, even though a more complete checking of the mandatory level data using increment, horizontal residual, and vertical residual checks would make the siginficant level checking more complete and accurate. However, performing all these other checks would require neighboring data to be used.

Since the mandatory level heights and temperatures are already being checked, there is a reasonably firm framework in which to check the significant levels. This note describes the method developed for this checking. For this data also, the objective is to determine which temperatures are in error, and to correct those that can be corrected. As for mandatory levels, the checks are not being used to determine instrument errors (although for the significant levels, unlike mandatory levels, there is some sensitivity to these errors).

It is found that almost 90% of all data definitely determined to be in error can be corrected. (About 9% of the errors are actually a misinterpretation of the pressure level, the rest are temperature errors.) Approximately 11 corrections are made for each 12 hour data collection.

The rest of this note will describe the checks that are used for significant level data, give examples of errors and corrections, and give statistics on the performance of the procedure.

The Method.

The development of a method to identify and correct bad significant level temperatures involved designing sensitive and independent (or as independent as possible) checks. There are two main kinds of checks that seem applicable: hydrostatic checks and interpolation checks. With the assumption of correct mandatory level data, the hydrostatic residual, using only mandatory level data, should be small. A hydrostatic residual is defined by

$$s_{i,i+1} = z_{i+1} - z_i - A_{i,i+1} - B_{i,i+1}(T_i + T_{i+1}), \qquad (1)$$

where

$$A_{i,i+1} = (RT_o/g)\ \ln(p_i/p_{i+1})$$

$$B_{i,i+1} = (R/2g)\ \ln(p_i/p_{i+1})$$

and

$$T_o = 273.15 \text{ Kelvin.}$$

If all data, including significant level temperatures, are used to calculate the hydrostatic residual, the value should generally be smaller than if mandatory levels alone are used. Fig. 1 shows the distribution of hydrostatic residuals for the lowest 6 mandatory layers, using only data at the mandatory levels. The distributions appear roughly Gaussian, with a mean that is near zero but slightly positive for the lowest layers, indicating the neglected influence of moisture on the thickness. Fig. 2 shows the distribution of hydrostatic residuals for the lowest 6 mandatory layers, but now including the significant levels.

Again, the distributions appear Gaussian and the means are even smaller. These features are emphasized in Fig. 3 which shows the mean and standard deviation of values of the hydrostatic residuals, using all data or using only mandatory level data. There is a small but consistent reduction to both statistics when the significant level data are used. The amount of the reduction depends to some extent upon the average curvature of the temperature profile and thus the required number of significant levels between any two mandatory levels.

The magnitude of the standard deviation of the hydrostatic residuals for a large sample of rawinsondes is used to specify a critical value, equal to about 7 standard deviations. Actual residuals larger in absolute value than these critical values, called admissible residuals, are considered large. The values used for each mandatory layer are shown in Fig. 4. The values shown for mandatory levels only are the ones used in the hydrostatic checking of mandatory level heights and temperatures.

When the hydrostatic residual, using only mandatory levels, is small, but the residual using all levels is large, then an error is suspected at (one of) the intervening significant level(s). It is possible to calculate what the value of every significant level temperature must be in turn to make the residual vanish, under the assumption that a single error is responsible for the large residual. The necessary adjustment to the significant level temperature, T', is calculated from

$$T'_j = T^*_j - T_j, \qquad (2)$$

where T_j is the reported temperature and

$$T^{*}_{j} = (B_{j-1,j} + B_{j,j+1})^{-1} \left[z_n - z_m - \sum_{i=m}^{n} A_{i,i+1} \right. \qquad (3)$$

$$\left. - T_m B_{m,m+1} - T_n B_{n-1,n} - \sum_{\substack{i=m+1 \\ i \neq j}}^{n-1} T_i (B_{i-1,i} + B_{i,i+1}) \right],$$

$$j = m+1, m+2, \ldots, n-2, n-1.$$

The mandatory levels surrounding the significant levels have vertical indices m and n. A value of T' is calculated for each of the intermediate significant levels. The value of T' gives one estimate of the error, once a value is definitely identified as wrong.

The second kind of checks is interpolation checks. Even though a significant level temperature is generally defined specifically because the temperature profile is not linear with respect to the logarithm of pressure, nevertheless, the temperature generally deviates not too greatly between adjacent significant levels from linearity. Three kinds of interpolation residuals are defined, all being linear with respect to the logarithm of pressure. The first is just the difference between the value and the linear interpolation from the nearest neighbors, whether they be significant or mandatory levels. If only one value is possibly in error, then this is the most appropriate interpolation check, but to avoid the confusion of the influence of bad neighbors, a second check is defined. Its residual is the difference between the value and an interpolation from the nearest mandatory level neighbors (where the values are assumed to be accurate). The third check forms the residual which is the difference between the temperature value with T' added and the interpolated value from the nearest neighbors. Fig. 5 shows graphically all these residuals.

Just as it is necessary to define when one of the hydrostatic residuals is large, so it is also necessary to define when the interpolation residuals are large. Experience with the hydrostatic quality control has shown that it is difficult to consider any temperature corrections of less that 10 degrees. For that reason, large interpolation error is defined as one of absolute value greater that 10 degrees. Testing has shown this value to be appropriate.

It will be useful to identify the various residuals by simple names. Table 1 gives those names. Fig. 5 graphically shows these various residuals. The hydrostatic residuals are valid only between mandatory levels, while the interpolation residuals are only valid at significant levels. The closed circles represent the observed temperatures, the open triangle represents the true temperature, and the open circle represents the temperature with T' added. The magnitude of T' is shown by the arrow below.

There is a pattern of large/small residuals that is indicative of definite temperature errors that may be confidently corrected. This pattern is illustrated by the example of Fig. 5: the residuals HYALL (by assumption), INTALL, and INTMND, as well as T' are large, while HYMND (by assumption) and INTTP are small. This pattern not only isolates the problem to be at a significant level, but it identifies the particular level, and in addition, two reasonable estimates of the error are generated. First, there is the value of T', and secondly, there is the interpolated residual INTALL. In practice, a weighted average of these two estimates is used to provide a provisional correction. Some further details of the error pattern diagnosis can be found in Appendix A. One feature of the temperature corrections is that in all cases a "simple" correction is searched for. A "simple" correction is one that is as close as possible to the suggested correction, but one which does one or more of the following: 1) modifies one digit, 2) changes the sign, or 3) rearranges the digits. About 2/3 of the corrections are "simple". Fig. 6 shows the distribution of error types for 17 cases in January and February 1990.

Pressure Corrections.

During the development of the method for correction of significant level temperatures, it was occasionally found that if the pressure were divided by 10, then the reported temperature fit well with the surrounding temperatures. Therefore, a modification was made to check for an error of this type before accepting any temperature correction. This error arises since the radiosonde report is entered onto communication lines in parts, with the data below 100mb separated from the data at and above 100mb. It is therefore possible to sometimes interpret the data as belonging to the incorrect part.

The present procedure checks each temperature correction first to make sure that it should not rather be a pressure correction, thus the procedure will only find pressure errors that would otherwise lead to temperature corrections. There are two ways in which the procedure will be improved in the future: 1) when a pressure error is found, correct all pressures for that part of the report, and 2) specifically look for pressure errors, even when no confident temperature correction is diagnosed. By including these two improvements, there would likely be several times more pressure errors corrected.

Examples of Corrections.

The following examples illustrate the most usual types of corrections: changes to a single digit, changes of sign, and changes of sign plus a single digit. These are examples of "simple" correction. The terminology may be seen in Fig. 5. The values in the examples are shown after the correction has been made.

Example 1. Correction to single digit.

Station: 12374 Date: 00Z 3 Nov 1990
P: 290. T_{old}: -33.3 T_{new}: -53.3

P	Z	T	T'	HYALL	HYMND	INTALL	INTMND	INTTP	IMAND	ISIG
300	8820	-51.7	-	-	-	4.0	-	-	0	0
290	-	-53.3	-20.0	-20.0	-	-19.1	-18.5	2.2	0	14
266	-	-54.1	-18.1	-18.1	-	9.9	2.1	18.7	0	9
250	9990	-52.1	-	-	-11.6	0.0	-	-	2	0

The Example 2 illustrates another feature of the significant level checking: a level at which the temperature is present, but the height is missing, is treated as a significant level. In the example, the sign of the temperature at 500mb is corrected.

Example 2. Correction of sign.

Station: 51777 Date: 00Z 3 Nov 1990
P: 500. T_{old}: 15.0 T_{new}: -15.0

P	Z	T	T'	HYALL	HYMND	INTALL	INTMND	INTTP	IMAND	ISIG
700	3098	4.4	-	-	-	7.8	-	-	0	0
500	-	-15.0	-30.0	-	-	-30.7	-30.7	0.6	0	14
400	7380	-29.1	-	-244.9	7.1	17.9	-	-	2	0

Example 3 combines corrections to sign and a single digit of the temperature. In all the examples, the pattern of IMAND and ISIG is the same: IMAND is 2 for the layer, indicating a problem, and ISIG is 14 at the level of confident correction, indicating an error at this very level, which is corrected sufficiently by adding T' to T for the level.

Example 3. Correction to sign and single digit.

Station: 94294 Date: 00Z 3 Nov 1990
P: 336. T_{old}: 18.8 T_{new}: -28.8

P	Z	T	T'	HYALL	HYMND	INTALL	INTMND	INTTP	IMAND	ISIG
400	7570	-17.5	-	-	-	12.2	-	-	0	0
336	-	-28.8	-47.6	-	-	-46.0	-46.0	-0.5	0	14
300	9660	-33.5	-	-190.6	3.2	15.0	-	-	2	0

Statistics of Operation.

The correction of significant level temperatures and pressures has been in operation since April 11, 1990, producing about 600 confident temperature corrections and 100 pressure corrections each month. The geographical distribution of the errors is shown in Figs. 7,8. Fig. 7 shows the total of all suspected significant level errors for 17-30 September, 1990, divided into geographical regions, each containing one or more WMO station blocks. The abbreviations are shown in Table 2. The number of suspected errors includes any instance where one or more of the residuals was large, therefore including many cases where there was either no error or where there is not sufficient information to determine the true nature of the error or to correct it. Fig. 8, on the other hand, shows only the numbers of errors that were actually corrected, for these same regions and time period. The relative distribution of these error corrections by region is similar to the distribution of mandatory level corrections, indicating a common cause. (See Office Note 369.) The fact that the overwhelming majority of corrections is "simple" indicates that the cause is human error.

Summary

This note has described one element in the development of a Complex Quality Control system at the National Meteorological Center: the checking of significant level temperatures. They are checked, using vertical interpolation and hydrostatic checks, and the analysis of possible errors is made in a Decision Making Algorithm. The checking of significant level temperatures follows the checking of mandatory level heights and temperatures, using the principle that the most accurate data, with the most powerful checks, be corrected first.

The interpolation and hydrostatic checks has been described in some detail; their complex can be used to diagnose confident corrections with nearly the same accuracy as the changes at mandatory levels. In fact, the greatest difficulty comes when the significant level checks are confused by the fact that an error has not been corrected at a mandatory level. However, in these cases, no correction is attempted. The corrections have been found to be "simple" in the majority of cases, showing the human element in the formation of the errors.

The performance of the algorithm has been demonstrated with statistics from half a month's operation. The geographical distribution of errors approximates that at mandatory levels, showing a common cause.

The next step in the development of the CQC is the checking of radiosonde heights and temperatures, using more complete checks: horizontal interpolation, vertical interpolation, and increment checks. As explained earlier, it would have been preferable to implement these checks before the significant level temperature checks, since then the mandatory level data would be more correct as a backbone for the significant level checking. With the introduction of these additional checks, however, there will be no changes necessary to the significant level checks--they will merely perform better. Following this step, or along with it, will be the development of checks of the radiosonde wind components. This development order is natural since each step will use data that are already checked.

Acknowledgements.

I would like to especially thank Lev Gandin for many valuable discussions during the development of the procedure. I would like to also thank Lauren Morone for her continuing monitoring of the corrections and her suggestions for improvement.

REFERENCES

Collins, W.G. and L.S. Gandin, 1990: Comprehensive hydrostatic quality control at the National Meteorological Center. *Mon. Wea. Rev.*,**118**,??.(to be published)

Gandin, L.S. and W.G. Collins, 1990: One year of operational comprehensive hydrostatic quality control at the National Meteorological Center, Office Note 369. [Available from Director, NMC, Washington, D.C. 20233.]

APPENDIX A

A method has been developed to systematize the description of the pattern of residuals by assigning two binary integers to the results of the various checks. The residuals HYALL, HYMND, and T' INTALL, INTMND, INTTP are either small or large, signified by a 0 or 1. The first two, HYALL and HYMND are combined in a binary integer, called IMAND, having a value of 0, 1, 2, or 3 (binary 00, 01, 10, or 11). Similarly, the interpolation residuals are combined into a binary integer (0000 through 1111), the binary digits standing for large/small for T', INTALL, INTMND, and INTTP. An "error type" is formed from the combination: "error type" = (-) (ISIG + IMAND). In this notation, a confident correction has IMAND = 2, ISIG = 14, so that the "error type" = -142. Table 3 shows the actual errors encountered for other "error types".

As an example, when ISIG = 15 and IMAND = 2, then all residuals and T' are large. There is a likely error at this level, but the suggested correction does not lead to good enough improvement so that the interpolation residual, REST, becomes small. The reason for the error is often undetermined, even by manual inspection. Another possibility in this case is that there is more than one error so that the value of T' obtained for any one layer is inappropriate; as a result, REST is large.

The most important case is for ISIG = 14 and IMAND = 2. This case was illustrated in Fig. 5, showing a single error with confident correction. Consider the residuals individually: HYMND is small since the mandatory level data are correct; HYALL is large since one significant level temperature is erroneous; T' is large since HYALL is large; INTMND and INTALL are both large since linear vertical interpolation from either nearest neighbors or nearest mandatory level neighbors will have a large residual; but INTTP is small since a combination of T' and linearly interpolated value will give a good starting point for a correction. Therefore, in this case there is a confident correction. As for hydrostatic corrections, a "simple" correction is attempted. As stated before, "simple" correction is one that is as close as possible to the suggested correction, but one which does one or more of the following: 1) modifies one digit, 2) changes the sign, or 3) rearranges the digits.

Table 1. Naming Convention for Residuals

HYALL	hydrostatic residual, using all levels
HYMND	hydrostatic residual, using only mandatory levels
INTALL	interpolation residual, using all levels
INTMND	interpolation residual, using only mandatory levels
INTTP	interpolation residual, using all levels, after T' is added to T.

Table 2. Large Regions

Abbr	Region Name	WMO blocks	Avg No.
W Eu	Western Europe	1-4,6-8,10,16	67.5
E Eu	Eastern Europe	9,11-13,15,17	24
USSR	USSR	20-38	177
W As	Western Asia	40-41	18
Ind	India, Ceylon	42-43	26
Mong	Mongolia	44	8
Jap	Hongkong,Taiwan,Korea,Japan	45-47	30
Indo	Indochina, Malaysia	48	8
Chin	China	50-59	123
Af	Northern & Central Africa	60-65,67	19.5
S Af	South Africa	68	11
USA	United States	70,72,74	87.5
Can	Canada	71	32
C Am	Central America	76,78	15.5
S Am	South America	80-87	16
Ant	Antarctica	89	9
P Is	Pacific Islands, Indonesia	91,96-98	24
Aust	New Zealand, Australia	93-94	24